Synthetic Evidential Study as Augmented Collective Thought Process -- Preliminary Report

Toyoaki Nishida[1], Masakazu Abe[1], Takashi Ookaki[1], Divesh Lala[1], Sutasinee Thovuttikul[1], Hengjie Song[1], Yasser Mohammad[1], Christian Nitschke[1], Yoshimasa Ohmoto[1], Atsushi Nakazawa[1], Takaaki Shochi[2,3], Jean-Luc Rouas[2], Aurelie Bugeau[2], Fabien Lotte[2], Ming Zuheng[2], Geoffrey Letournel[2], Marine Guerry[3], Dominique Fourer[2]

[1] Graduate School of Informatics, Kyoto University, Sakyo-ku, Kyoto, Japan
{nishida}@i.kyoto-u.ac.jp, {abe, ookaki,lala, thovutti,song}@ii.ist.i.kyoto-u.ac.jp, yasserfarouk@gmail.com, {christian.nitschke, ohmoto, nakazawa.atsushi}@ i.kyoto-u.ac.jp
[2] LaBRI, Bordeaux, France
{jean-luc.rouas, aurelie.bugeau}@labri.fr, fabien.lotte@inria.fr, {zming, geoffrey.letournel, fourer}@labri.fr
[3] CLLE-ERSS UMR5263 CNRS, Bordeaux, France
{takaaki.shochi, marine.guerry}@labri.fr

Abstract. Synthetic evidential study (SES) is a novel approach to understanding and augmenting collective thought process through substantiation by interactive media. It consists of a role-play game by participants, projecting the resulting play into a shared virtual space, critical discussions with mediated role-play, and componentization for reuse. We present the conceptual framework of SES, initial findings from a SES workshop, supporting technologies for SES, potential applications of SES, and future challenges.

Keywords. Group learning assistance, intelligent virtual agents, role play

1 Introduction

A collective thought process becomes more and more critical in the network age as a means for bringing together limited intelligence embodied by natural or artificial agents. A powerful methodology is needed to make collective thought processes effective. Methodologies such as brainstorming or mind map have been invented but they are mostly the third-person understanding and are limited in terms of actuality. Their output can only appeal to people through narratives or other pedagogical media. It is pretty hard for the ordinary audience to share the thought in terms of vivid and immersive understanding, or first-person understanding, of the output unless enough background knowledge is shared. The problem might be solved if it is presented as an interactive movie, but a huge cost would be required for that. Even with the existing state-of-the-art technology, however, it still appears beyond our scope to build a tool that allows for creating meaningful low-cost movies.

A less challenging, but still useful goal might be to build an intelligent tool that would allow people to progressively build a story base, which is a background setting consisting of pieces of story scenes, each of which consists of events played by one or more role actors with reference to the physical or abstract background. A story base may serve as a mother from which individual stories and games may be spawned. We assume that a story base will greatly help professional storytellers and game players produce high-quality content.

The long-term goal of this project is to establish a powerful method for allowing everybody to participate in a collective thought process for producing a story base for a given theme. There is a huge area of applications in education and entertainment. In addition, we believe that the project benefits science and technology. On the scientific hemisphere, it will open up a new methodology for investigating in-situ human behaviors. On the engineering hemisphere, it will significantly benefit not only content production but also product prototyping and evaluation.

In this paper, we portray a novel approach, called *synthetic evidential study* (SES), for understanding and augmenting collective thought process through substantiated thought by interactive media. The proposed approach draws on authors' previous work, including conversational informatics, human-computer interaction, computer vision, prosody analysis and neuro-cognitive science [1]. In what follows, we present the overview of SES, preliminary implementation of its components, and future perspectives.

2 Overview of SES

The present version of SES basically consists of four stages. The first stage is role-play by participants. Actors are invited to play a given role to demonstrate their first-person interpretation in a virtual space. A think aloud method is used so the audience can hear the background as well as the normal foreground speech. Each actor's behaviors are recorded using audio-visual means. The second stage is projecting role-play into a shared virtual space. The resulting theatrical play as an interpretation is recorded and reproduced for criticism by the actors themselves. The third stage is critical discussions with mediated role-play. It permits the participants or other audience to share the third-person interpretation played by the actors for criticism. The actors revise the virtual play until they are satisfied. The understanding of the given theme will be progressively deepened by repeatedly looking at embodied interpretation from the first- and third- person views. The final stage is componentization for reuse. The mediated play is decomposed into components and stored in the story base.

For illustration, suppose a handful people become interested in some scene of Romeo and Juliet by William Shakespeare. First, the participants will set up a SES workshop comprising the stages *a*-*c*, where the participants may either start from scratch or take up a previous piece of interpretation from the story base, criticize it, and produce their own, depending on their interest. Each one of them is asked to demonstrate her or his first-person interpretation for the scene, by following the events and expressing her or his thought as a behavior that she or he thinks the role would have acted for each event. Reproducing the behavior of a role, i.e. Romeo or Juliet, in each given scene will allow

the participant to feel the role's mental and emotional state, resulting in deeper and immersive understanding of the scene. On the second stage, they criticize with each other to improve the shared interpretation. The third person perspectives would allow the participants to gain the holistic understanding of the scene. Discussions permit the participants to know other possibilities of interpretation and their strength and weakness.

In order to maximally benefit from the above-mentioned aspects of SES, we need a powerful computational platform. It should be built on a distributed platform as participants would like to participate in from geographically distant points. We have found that the game engine, Unity 3D[1] in particular, best fits this purpose. It allows us to share a virtual space with complex objects and animated characters. Reproducing participants' theatrical role-play as Unity objects allows for implementing the stage *d* to progressively construct a story base for a community. It will enable its members to exploit the components of interpretation to build and share sophisticated knowledge about subjects of common interest.

The conceptual framework for the SES support technology consists of a shared virtual space technology for interfacing the users with the story world and the discussion space technology for supporting criticism and improving play as interpretation.

Virtual space technology plays a significant role to configure vast varieties of conversational environments. Consider the actors are asked to interpret the balcony scene of the Romeo and Juliet play. Although a physical setting for the balcony scene is critical for interpreting the details, it is too expensive and hence infeasible for ordinary cases unless the studio is available. Another difficulty arises when the number of participants is not enough. Whereas a virtual environment inhabited by non-playable characters solves difficulties in general, problems remain regarding how to generate qualified settings and non-playable characters for interpretation. Our technical contributions mostly address the second issue, while we utilize existing techniques for the first issue. Furthermore, the interface should be immersive and gesture-driven so participants can concentrate on the SES activities. Our technological supports for SES consist of the shared virtual space, virtual character realization, and discussion support.

We draw on technologies that have been developed in pursuit of conversational informatics. Our technology for SES not only supports and records conversations in a virtual immersive environment and projects the behaviors of the actors to those of synthetic characters but also analyzes interactions. It will help the participants deepen their interpretation even from the viewpoint of the second person or the interactant through interaction, which is only available by the virtual technology.

SES significantly extends the horizon of conventional pedagogy, as the participants of the SES workshops may be able to learn to view a phenomenon from multiple angles including not just active participants but also the perspective of other role players in real time if the agent technology is fully exploited. It should be extremely effective for social education, such as one for anti-bullying, as it allows the participants to "experiment" social affairs from different perspectives, which is almost impossible otherwise. From scientific points of view, the SES enables the understanding of human behaviors

[1] http://www.unity3d.com

in a vast variety of complex situations. As a result, one can design experiments far more realistic than conventional laboratory experiments. For example, the experimenters can slip into the SES session cues or distractors in a very natural fashion.

In what follows, we elaborate the SES workshop, the supporting technologies we have developed so far, and how SES is applied to conduct study human behaviors.

3 SES Workshop

The SES workshop is a joint activity open for every group of people to figure out a joint interpretation of a given theme by bringing together prior interpretations of participants. Repetition of acting together and discussion is a critical feature of a SES workshop[2]. In order to gain the practical features of the SES workshop, we conducted a preliminary workshop to gain initial insights about SES. We chose a story called Ushiwaka and Benkei[3] because it was very popular in Japan though the details are not well considered as it is rather a fiction told for children though it is partly based on historical fact. We had conducted one 1-hour session in which four participants repeated two discussion-play cycles. Due to the limitation of our measurement facility at that time, each one-role player acted for his part (Fig. 1a). The actor's motion was recorded by a Kinect[4] and projected as the behavior of a Unity agent. This is the topic of the subsequent discussions (Fig. 1b).

(a) Play and record

(b) Criticize and improve

Fig. 1. Snapshots from the preliminary SES workshop.

According to an informal *a posteriori* interview, the participants were able to be well involved in the discourse and obtained a certain degree of immersive understanding. In fact, we observed that the participants got interested in the details of the story such as how to handle a long sword. It motivated the participants to collect more information from the net and reflected that on their role play acts. In addition, the SES workshop

[2] In fact, a single player can conduct the SES workshop by leveraging the SES technology of virtual shared space and characters, if somebody prefers working alone.

[3] Ushiwaka and Benkei story goes like this. When Ushiwaka, a young successor to a noble Samurai family which once was influential but which was killed by the opponents, walked out of a temple in a mountain in the suburbs of Kyoto where he was confined, to wander around the city as daily practice, he met Benkei, a strong priest Samurai on the Gojo Bridge. Although Benkei tried to punish him as a result of having been provoked by a small kid Ushiwaka, he couldn't as Ushiwaka was so smart to avoid Benkei's attack. After a while, Benkei decided to become a life-long guard for Ushiwaka.

[4] http://www.xbox.com/Kinect

seems to work well for integrating partial knowledge of participants by acting and discussion together.

4 Computational Platform for SES

The conversation augmentation technology [1] supports immersive interactions made available by a 360-degree display and surround speakers and audio-visual sensors for measuring the user's behaviors (Fig. 2). The "cell" can be connected with each other or with other kinds of interfaces such as a robot so that the users can participate in interactions in a shared space.

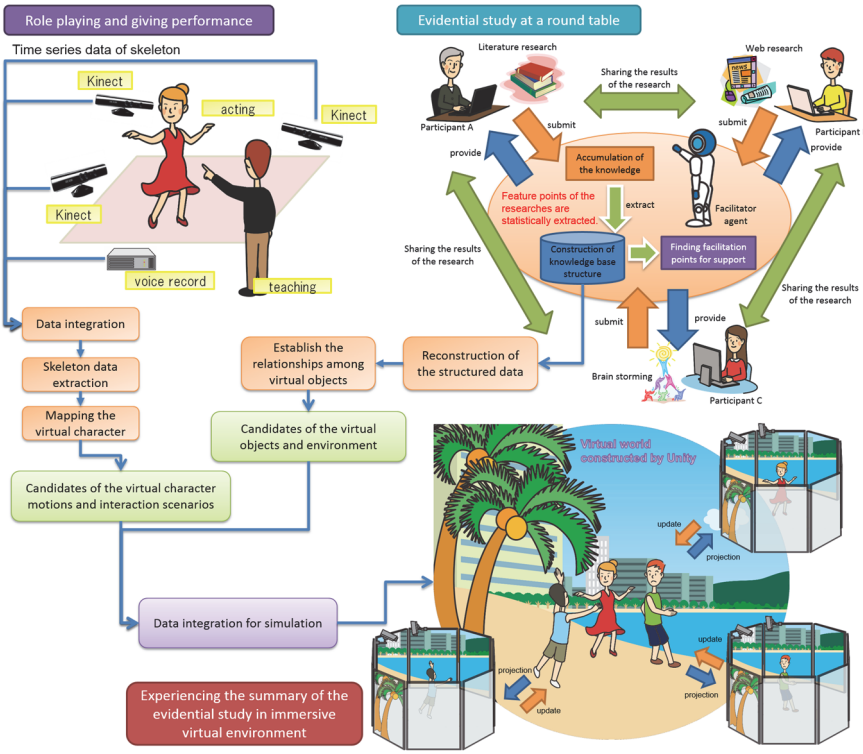

Fig. 2. The computational platform for SES.

It allows to project the behaviors of a human to those of an animated character who habits in a shared virtual space. The computational platform is coupled with the Unity platform so the participants can work together in a distributed environment.

A virtual space builder called FCWorld [2] enables the immersive environment to be linked to external software like Google Street View to benefit from various content available on the net. FCWorld was used to implement virtual network meetings with the Google Street View background.

5 Virtual Character Realization

Animated characters are critical elements of SES. Roughly speaking, a virtual character realization consists of two phases: character appearance generation and behavior generation.

As known in nonverbal communication, not only their behaviors but also their appearance may significantly influence the nature of interactions. We want to make the appearance of characters as unique as possible so they maximally reflect the target interpretation. An avatar with the face of an existing human might be expressive but too specific. Designing a game character is too expensive for ordinary SES sessions. We believe the de-identification approach [3] is the most effective. The idea is to take an existing human face and remove identifiable features while maximally retaining the expressive aspects. Ideally, voice de-identification should be coupled with visual de-identification.

Behavior generation is another key technical topic in virtual character realization. On the one hand, sophisticated behavior generation is necessary to express subtleties of mental state. On the other hand, participants should be easy and natural so they can concentrate on essential issues. Although recent advanced technologies using inexpensive depth-color sensors such as Kinect exist, there is plenty of room for improvement, such as full body behavior generation that integrates facial expression and bodily movement generation [1]. Although technologies are available for generating point cloud image representation of humans, more work needs to be done to accommodate point cloud representation into the game platform. Alternative approach is to employ crowd sourcing to collect typical behaviors that can serve as a prototype adaptable for a given purpose [4].

Learning by imitation is a powerful framework. So far, we have develop a basic platform and a more powerful imitation engine which monitors the behavior of the target continuously, autonomously detects recurrent signals and infers causality among observed events [5]. Imitation can play several roles at different stages of the SES session. For example, direct pose and motion copying can be used during role play to generate the preliminary behaviors that will be discussed and improved by the group. Our pose copying system is based on a modular decomposition of the problem that simplifies the extension to non-humanoid characters. During motion update, a combination of motion copying and the correction by repetition technology we developed earlier [6] can be used to improve the motions smoothly.

6 Discussion Support

Another critical phase of the SES support technology is discussion support. Our discussion support consists of the round-table support and the full-body interface. The ultimate goal of the former is to build a chairperson agent who can support discussions by estimating distribution of opinions, engagement, and emphasizing points. We have implemented several prototypes [7-9]. These prototypes focused on interactive decision-making during which people dynamically and interactively change the focusing

points. Our technologies allow to capture not only explicit social signals that clearly manifest on the surface but also tacit and ambiguous cues by integrating audio-visual and physiological sensing. As for the latter, we exploit Kinect technology to measure and criticize physical display of interpretation played together one or more local participants [1]. Since the system we developed is easy to setup measuring environment for human motion capture, we can conduct SES anywhere indoors.

7 First-Person View by Corneal Imaging

Capturing the first person view of the world provides a valuable means for estimating the mental status of a human either in a role-playing game or in discussion. In fact, we have found that first-person view may bring about quite different emotional state from the third-person view in human-robot interaction [10]. By exploiting the fact that the cornea of a human reflects the surrounding scene over a wide field of view, our corneal imaging technology allows for determining the point of gaze (PoG) and estimating the visual field from reflecting light at the corneal surface using a closed-form solution. Compared to the existing approaches, our method achieves equipment and calibration-free (PoG), depth-varying environment information and peripheral vision estimation. In particular, the first and the third are very important to be used as human interface devices and beyond the current human view understandings that uses only the 'point' of the gaze information [11].

8 Prosody Analysis

Prosody analysis is a key technology for estimating and distinguishing social affects, such as laughter and smile, in face-to-face communication. Production and interpretation of social affects is indispensable for obtaining in-depth understanding of the SES sessions. So far, our prosody analysis and corpus-building technologies have been applied to analysis of laughter/smile/sad speech and cross-cultural communication. Intended affective meanings are conveyed by various modalities such as body movements, gestures, facial expressions as well as vocal expressions. In particular, vocal expressiveness of these affects is intensively studied since the 90's. Recently, Riliard et al. envisage the prosodic variations, which are used to encode such social affects. They also try to identify the characteristics of these prosodic codes in competition with others (e.g. syntactical and lexical prosodic configuration) [14]. Such prosodic analysis focusing on social affective meaning may be implemented by cross-cultural communication processing. Thanks to this methodological approach some universal and culture-specific prosodic patterns were identified even for the same label of affect (e.g. surprise).

Combined with the SES technology, our prosody analysis techniques will be extended to multi-modal prosody analysis, powerful enough to investigate complex social-emotional communication. Coupled with brain activity visualization technique such as [13], it will allow us to investigate brain activities underlying the social affects.

9 SES-based Human Behavioral Science

SES can serve as a novel platform for conducting human behavioral study as it allows the researchers to build a sophisticated experiment environment.

(1) Building multi-modal corpus for cross-cultural communication. Socio-emotional aspects are critical in understanding cross-cultural communication. Other cross cultural studies on affective speech are conducted in cross cultural paradigm among four languages: Japanese, American English, Brazilian Portuguese and French [14,15]. The results showed that subjects of different cultural origins shared about 60 % of the global representation of these expressions, that 8% are unique to modalities, while 3 % are unique to language background. The results indicate that even if specific cultural details punctually play an important role in the affective interpretations, it may also emphasize the fact that, a great deal of information is already shared amongst speakers of different linguistic backgrounds.

Therefore, the SES platform permits researchers to set up complex situations to quantitatively investigate socio-emotional aspects of cross-cultural communication in the shared virtual space so that culture-dependent cues can be observed and contrasted in multiple desired conditions, as a natural extension to our method of analyzing prosodic aspects.

(2) Dance lesson. Measure and criticize the detailed quantitative aspects of motion. Coupled with group annotation tools, we can analyze the detailed nonverbal behaviors of tutor-student interactions. The group annotation tools are useful for building consensus because they automatically extract and propose feature points for criticizing. The scheme can be used to analyze how people criticize the features of played actions [1].

(3) Laughter. In this study, we mostly depend on physiological sensors to make subtle distinctions among hidden laughter, enforced laughter, and genuine laughter. Especially hidden laughter is an important cue to get the true response. The SES framework most fits more comprehensive study on comedic plays that induce laughter on the audience [16].

(4) Virtual basketball. In this study, we shed light on how social signals from beginning players and non-playable characters in motion can be used to read intentions from friends and opponents. In the future, the study may be extended to highlight the group discussions for criticizing and improving performance [17].

(5) Cultural crowd. This study aims at gaining first-person understanding of crowd in a different culture. Currently, we are focusing on queuing behaviors in a different culture, trying to identify social signals and norm in a given culture, by conducting a contrastive study [18].

(6) Physiological study and evaluation of social interactions in the virtual space. Being able to assess how observers perceive, from an affective (e.g., emotions) and cognitive (e.g., attention, engagement) point of view, the play of the actors, could provide very interesting insights to refine and improve the play and/or the story. Similarly measuring such affective or cognitive states during collaborative tasks performed in the joint virtual space could be used to study, assess and then optimize the collaborative work. Interestingly enough, we and others have shown that such affective and cognitive states could be measured and estimated in brain signals (electroencephalography) [19] or in

other physiological signals (heat rate variability, galvanic skin response, etc.) [20]. The SES would thus provide a unique test bed to study physiological based assessment of both affective and cognitive experience and distance collaboration tasks.

10 Concluding Remarks

In this paper, we introduced synthetic evidential study (SES) as a novel approach for understanding and augmenting collective thought process. We presented the conceptual framework of SES, initial findings from a SES workshop, supporting technologies for SES, and potential applications of SES. We believe that SES has many applications ranging from science to engineering, such as content production, collaborative learning and complex human behavior analysis. Future challenges include, among others, virtual studio, layered analysis of human behavior, and multi-modal prosody analysis.

The SES paradigm has opened up numerous new challenges as well as opportunities. Among others, we have recognized three challenges as the most useful to enhance the current framework of SES. The first is virtual studio that permits user to set up and modify a complex terrain and objects on the fly. The second is analysis of participants' action and separation of layers, as a participant' behavior refers to either the action of the role or meta-level action such as a comment to her or his action. The third is a methodology for guiding SES sessions to bring about intended effects on the participants, depending on the purpose. For example, if the purpose of a given SES activity is to help the participants design a new industrial product, it should be very helpful if the participants are encouraged to consider potential product usage scenarios in a comprehensive fashion.

Acknowledgments

This study has been carried out with financial support from the Center of Innovation Program from Japan Science and Technology Agency, JST and AFOSR/AOARD Grant No. FA2386-14-1-0005, JSPS KAKENHI Grant Number 24240023, the French State, managed by the French National Research Agency (ANR) in the frame of the "Investments for the future" Programme IdEx Bordeaux (ANR-10-IDEX-03-02), Cluster of excellence CPU. We are grateful for Peter Horsefield who helped us improve the presentation of this paper.

References

1. Nishida, T., Nakazawa, A., Ohmoto, Y., Mohammad, Y.: Conversational Informatics–A Data-Intensive Approach with Emphasis on Nonverbal Communication, Springer (2014)
2. Lala, D., Nitschke, C., Nishida, T.: Enhancing Communication through Distributed Mixed Reality. in: Proc. AMT 2014: 501-512 (2014)
3. Letournel, G., Bugeau, A., Ta, V.-T., Domenger, J.-P., Gallo, M. C. M.: Anonymisation fine de visages avec préservation des expressions faciales, Reconnaissance de Formes et Intelligence Artificielle (RFIA) 2014, Rouen : France (2014)

4. Han, X., Zhou, W., Jiang, X., Song, H., Zhong, M., Nishida, T.: Utilizing URLs Position to Estimate Intrinsic Query-URL Relevance, in: Proc. ICDM 2013, pp.251-260, 2013 (2013)
5. Mohammad, Y., Nishida, T.: Robust Learning from Demonstrations using Multidimensional SAX, to be presented at 14th International Conference on Control, Automation and Systems (ICCAS 2014), Gyeonggi-do, Korea (2014)
6. Mohammad, Y., Nishida, T.: NaturalDraw: Interactive Perception Based Drawing for Everyone. in: Proc. IUI 2007, pp: 251-260 (2007)
7. Ohmoto, Y., Miyake, T., Nishida, T.: Dynamic estimation of emphasizing points for user satisfaction evaluations. In Proc. the 34th annual conference of the cognitive science society pp. 2115–2120 (2012)
8. Ohmoto, Y., Kataoka, M., Nishida, T.: Extended methods to dynamically estimate emphasizing points for group decision-making and their evaluation. Procedia-Social and Behavioral Sciences, 97, 147–155 (2013)
9. Ohmoto, Y., Kataoka, M., & Nishida, T.: The effect of convergent interaction using subjective opinions in the decision-making process. In Proc. the 36th annual conference of the cognitive science society, pp. 2711-2716 (2014)
10. Mohammad, Y., Nishida, T.: Why should we imitate robots? Effect of back imitation on judgment of imitative skill (submitted).
11. Nitschke, C., Nakazawa, A., Nishida, T.:I See What You See: Point of Gaze Estimation from Corneal Images, Proc. 2nd IAPR Asian Conference on Pattern Recognition (ACPR), pp.298-304 (2013)
12. Rilliard, A., De Moraes, J., Erickson, D., Shochi, T.: Social affect production and perception across languages and cultures - the role of prosody. Leitura, 52 (forthcoming).
13. J. Frey, R. Gervais, S. Fleck, F. Lotte, M. Hachet, Teegi: Tangible EEG Interface, ACM User Interface Software and Technology (UIST) symposium (2014)
14. Rilliard, A., Erickson, D., De Moraes, J., Shochi, T.: Linguistic approaches to emotions, in context, chapitre Cross-Cultural Perception of some Japanese Expressions of Politeness and Impoliteness, pages 251–276. John Benjamins, Amsterdam (2014)
15. Fourer, D., Shochi, T., Rouas, J.-L., Aucouturier, J.-J., Guerry, M.: Prosodic analysis of spoken Japanese attitudes, in: Proc. Speech Prosody 7, pp. 149--153 (2014)
16. Tatsumi, S., Mohammad, Y., Ohmoto, Y., Nishida, T.: Detection of Hidden Laughter for Human-agent Interaction. Procedia Computer Science, 35, pp. 1053-1062 (2014)
17. Lala, D., Mohammad, Y. Nishida, T.: A joint activity theory analysis of body interactions in multiplayer virtual basketball, 28th British Human Computer Interaction Conference, Southport, UK (2014)
18. Thovuttikul, S., Lala, D., van Kleef, N., Ohmoto, Y., Nishida, T.: Comparing People's Preference on Culture-Dependent Queuing Behaviors in a Simulated Crowd, in: Proc. ICCI*CC 2012, pp. 153-162 (2012)
19. Frey, J., Mühl, C., Lotte, F., Hachet, M.: Review of the use of electroencephalography as an evaluation method for human-computer interaction, International conference on Physiological Computing Systems (PhyCS 2014), pp. 214-223 (2014)
20. Fairclough, S., Fundamentals of physiological computing, Interacting with Computers, 2009, 21, 133-145 (2009)